yang_seonhyeHW10

Seonhye Yang

2/12/2019

```
library(data.table)
hybrid_reg <- fread("http://users.stat.ufl.edu/~winner/data/hybrid_reg.csv")
hybrid <- data.table("hybrid_reg.csv", header=T)
```

Question 1

```
attach(hybrid_reg, warn.conflicts = F)
model<- lm(msrp~mpgmpge+accelrate)

modelhat<-hatvalues(model)
which.max(modelhat)
```

```
## 151
## 151
```

```
max(modelhat)
```

```
## [1] 0.09304655
```

```
lsr<-residuals(model)
which.max(lsr)
```

```
## 37
## 37
```

```
max(lsr)
```

```
## [1] 51093.28
```

```
rstandard<-rstandard(model)
which.max(rstandard)
```

```
## 37
## 37
```

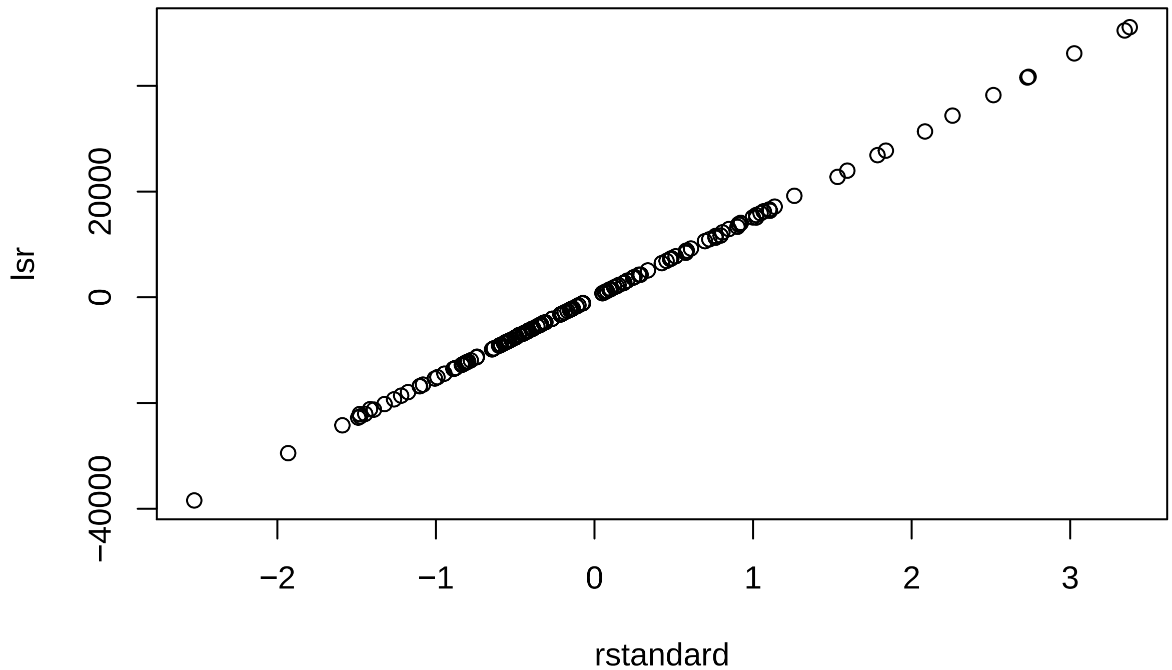
```
max(rstandard)
```

```
## [1] 3.375688
```

151 has the largest leverage with a measure of 0.09304655. 37 has the largest least squares residual with a value of 51093.28. 37 has the largest studentized residual with value 3.375688.

Question 2

```
plot(rstandard, lsr)
```



Leverage is how far away the independent variable values of an observation are from those of the other observations. Since the leverages for the variables are different, this might explain why not all the points are on the same line. The studentized residuals are not linear to leastsquares residuals because of the different leverages variables have. ## Question 3 Show that $\sum_{i=1}^{n} h_{ii} = tr(H) = p\prime$

$$A^{-1}A = I$$

We already know that

$$H = X(X^TX)^{-1}X^T$$

where

$$A = X$$

and

$$B = (X^TX)^{-1}X^T$$

so

$$tr(AB) = tr(X(X^TX)^{-1}X^T)$$

We also know that the trace

$$tr(A) = \sum_{i=1}^{n} a_{ii}$$

and

$$tr(AB) = \sum_{i}^{n} (ab)_{ii}$$

Since

$$X(X^TX)^{-1}X^T = (X^TX)^{-1}X^TX$$

this will equal to an identity matrix, I

Therefore,

$$tr(I_{p\times p}) = p\prime$$

Question 4

$$E(\hat{\epsilon_i}^2) = E(\hat{\epsilon_i}\epsilon^T)$$

We also know that

$$\hat{\epsilon} = y - \hat{y} = (I - H)y = (I - H)(y - X\beta)$$

so:

$$E(\hat{\epsilon_i}\epsilon^T) = (I - H)(y - X\beta)(I - H)(y - X\beta)^T(I - H)$$
$$= (I - H)E(y - X\beta)(y - X\beta)^T(I - H)(I - H)$$

$$= \sigma^2(I - H)$$

So therfore, $E(\hat{\epsilon_i}^2) = \sigma(I - H)$

Question 5

We need to show that $E(s^2) = \sigma^2\backslash$ We know that $E(\hat{\epsilon}_i^2 = \sigma^2(\mathbf{I} - h_{ii})$ and that $s^2 = \frac{1}{n-1}\Sigma_{i=1}^n(x_i - \bar{x})^2\backslash$ So

$$E(s^2) = E(\frac{1}{n-1}\Sigma_{i=1}^n(x_i - \bar{x})^2)$$

$$= \frac{1}{n-1}E(\Sigma_{i=1}^n(x_i - \bar{x})^2)$$

We know that $E(\Sigma_{i=1}^n(x_i - \bar{x})^2) = \sigma^2(\mathbf{I} - h_{ii})$ from earlier, so we get:

$$= \frac{(n-1)\sigma^2}{(n-1)}$$

$$E(s^2) = \sigma^2$$

This shows that σ^2 is an unbiased estimator.